Supplemental Data Table of Contents

Data Title	Data Description	Data available in File
Method S1: IGM Bioinformatics Processing	IGM Bioinformatics Processing	CKD collapsing JASN SuppMat_Final.pdf
Method S2: Quality Metrics	Quality Metrics used in Collapsing Analyses	As above
Method S3: Missense intolerance scores	Missense intolerance scores description	As above
Method S3: UK biobank Lookup	Top genes from the collapsing analysis and CKD phenotypes in a UK biobank analysis	As above
Figure S1: *APOL1* PCA plots	PCA plots of the *APOL1* cases and population-matched controls	As above
Figure S2: QQ plots of synonymous variants	QQ plots of synonymous variants (experimental negative control) in each sub-group	As above
Figure S3: *CPT2* enrichment, MTR filter	Lollipop plots and bar chart of *CPT2* enrichment using the MTR <= 50% filter	As above
Figure S4: Power curve	Power curve for different control carrier frequencies	As above
Table S1: Cohort Demographics	Cohort Demographics	As above
Table S2: Control cohort phenotypes	Control cohort phenotypes	As above
Table S3: Qualifying Variant parameters and subgroups	Qualifying Variant parameters and subgroups	As above
Table S4: LIMBR and MTR comparison	LIMBR and MTR comparison to standard missense-only collapsing analysis run	As above
Table S5: All genes across all models where P < 0.05	All genes across all models where P < 0.05	SUPP_TABLE_S5_Final.xlsx
Column descriptors for Table S5	Column descriptors for Table S5	CKD collapsing AJHG SuppMat_Final.pdf
Table S6: Clinical phenotypes in known genes	Distribution of phenotypes in cases with qualifying variants in PKD1, PKD2, COL4A3, COL4A4 and COL4A5	As above

Table S7: Ancestry prediction in modifier analysis	Ancestry predictions of cases in modifier analysis groups	As above
Table S8: *AHDC1* rare QV rates in *APOL1*	*AHDC1* rare QV rates in cases with *APOL1* genotypes against ancestry-matched controls	As above
Table S9: Look-up of UK biobank data	Look up of top genes in a recent analysis of UK biobank data	As above
Table S10: Predicting p-value signals	Predicting possible p-value signals among an increasing sized test cohort	As above

1. Supplemental Method: IGM Bioinformatics Processing

Samples were evaluated first across 12,840 single nucleotide polymorphisms (SNPs). These SNPs were previously shown in-house to be informative across diverse human populations, and reliable for genotyping exome sequence, and so were ideal for identifying cryptic relationships and population stratification. The exomes of a full cohort of 3,169 CKD patients and 9,996 controls were assessed for cryptic relatedness using KING[1]. Where pairs of samples were estimated as second-degree relatives or nearer, one of the pair was removed, favouring the inclusion of cases over controls. All samples were then compared across 10 principal components (PCs) using EIGENSTRAT[2] and outliers (±6 standard deviations across any PC) were excluded. Following these QC steps a total of 3,150 cases and 9,563 controls remained.

2. Supplemental Method: Quality Metrics used in Collapsing Analyses

In addition to the criteria described in the Methods section, we required variants to have: i) at least 10-fold coverage, ii) quality score (QUAL) of at least 50, iii) genotype quality score (GQ) of at least 20, iv) quality by depth score (QD) of at least 2, v) mapping quality score (MQ) of at least 40, vi) read position rank sum score (RPRS) greater than -4, vii) mapping quality rank sum score (MQRS) greater than -10, viii) Fisher's strand bias score (FS) less than 200 (indels) or less than 60 (SNVs), ix) for heterozygous genotypes, the alternate allele fraction was required to be at least 0.25. We also excluded variants if they were among a predefined list of known sequencing artefacts (a list of previously established in-house of variants showing a significant frequency imbalance between European controls in our sequencing database and European samples in EVS despite good coverage) or if they were marked as being problematic variants by EVS as failed with their SVM classification (http://evs.gs.washington.edu/EVS/HelpDescriptions.jsp) or ExAC[3] (using VQSLOD filter).

3. Supplemental Method: Missense intolerance scores

We adopt two novel approaches to identifying missense intolerant regions within genes. The sliding-window Missense Tolerance Ratio (MTR) quantifies missense variant intolerance within a gene independent of known functional boundaries.[4] This metric quantifies how missense depleted a 31-codon window of a gene is based on the gnomAD exome standing variation data. Comparing ClinVar pathogenic variants to novel control missense variants shows striking enrichment for pathogenic variants occurring in the missense intolerant sub-regions of protein-coding genes (http://mtr-viewer.mdhs.unimelb.edu.au/). The localized intolerance model using Bayesian regression (LIMBR), extends the work in our earlier subRVIS regional collapsing[5] by fitting a Bayesian multi-level model to gene exons according to the Conserved Domain Database[6, 7], regressing the number of missense variants on the total number of variants to leverage information about how missense variants within the same gene or gene sub-region may be related to each other.

To perform regional missense-intolerance informed collapsing analyses, similar to the use of the PolyPhen-2 score[8], we included missense-specific filters to further require that a missense variant also occur in an missense intolerant region based on either the MTR or LIMBR. For MTR we required missense variants occur among the 50th percentile most missense intolerant sliding windows specific to that gene. It considers case/control missense variants occurring in the tolerant 50% of protein-coding sliding windows of a gene as no longer qualifying. For LIMBR, we selected a filtering threshold of the 50th percentile for our collapsing analysis. Missense variants in exons with LIMBR percentiles greater than 50 were excluded from analysis [9].

4. Supplementary methods. UK biobank Lookup

We selected all kidney associated traits from the UK Biobank (cite Canela-Xandri et al., Nature genetics 2018) data and downloaded their GWAS p-values for genotyped and imputed variants from GeneATLAS (http://geneatlas.roslin.ed.ac.uk/, last accessed 2018-12-01). Quality control for imputed variants included ensuring an imputation score ≥ 0.9, a MAF of ≥ 0.001, and a Hardy Weinberg Equilibrium p-value of ≥ 10^{-10}. All variants with a p-value of < 10^{-4} for any of the kidney traits were included as variants with suggestive GWAS signal for kidney disorders. Finally, we mapped the variants to genes by including 50 kb up- and downstream of the gene. The final list consisted of 7,542 out of 18,852 genes.

List of kidney related traits in UK Biobank

key	**Description**
clinical_c_Block_N00-N08	N00-N08 Glomerular diseases
clinical_c_Block_N10-N16	N10-N16 Renal tubulo-interstitial diseases
clinical_c_Block_N17-N19	N17-N19 Renal failure
clinical_c_Block_N20-N23	N20-N23 Urolithiasis
clinical_c_Block_N25-N29	N25-N29 Other disorders of kidney and ureter
clinical_c_Block_O10-O16	O10-O16 Oedema, proteinuria and hypertensive disorders in pregnancy, childbirth and the puerperium
clinical_c_I12	I12 Hypertensive renal disease
clinical_c_N03	N03 Chronic nephritic syndrome
clinical_c_N12	N12 Tubulo-interstitial nephritis, not specified as acute or chronic
clinical_c_N13	N13 Obstructive and reflux uropathy
clinical_c_N17	N17 Acute renal failure
clinical_c_N18	N18 Chronic renal failure
clinical_c_N19	N19 Unspecified renal failure
clinical_c_N20	N20 Calculus of kidney and ureter
clinical_c_N23	N23 Unspecified renal colic
clinical_c_N28	N28 Other disorders of kidney and ureter, not elsewhere classified
selfReported_n_1213	renal/kidney failure
selfReported_n_1217	urinary tract infection/kidney infection
selfReported_n_1218	kidney stone/ureter stone/bladder stone
selfReported_n_1454	other renal/kidney problem

Figure S1 Ancestry Matching for *APOL1* analysis

Filters for calling *APOL1* G1 / G2 variants: Min base coverage: 3. Min QUAL (GATK): 30. Var-status (GATK): pass or intermediate

Control and non-dual-risk cases selection

Figure S1A: PCA plot for *APOL1* dual-risk cases (cases are homozygous or compound het for G1/2 alleles)

Figure S1B: PCA plot for no APOL1 dual risk cases (cases are heterozygous G1/2 or homozygous reference for both).

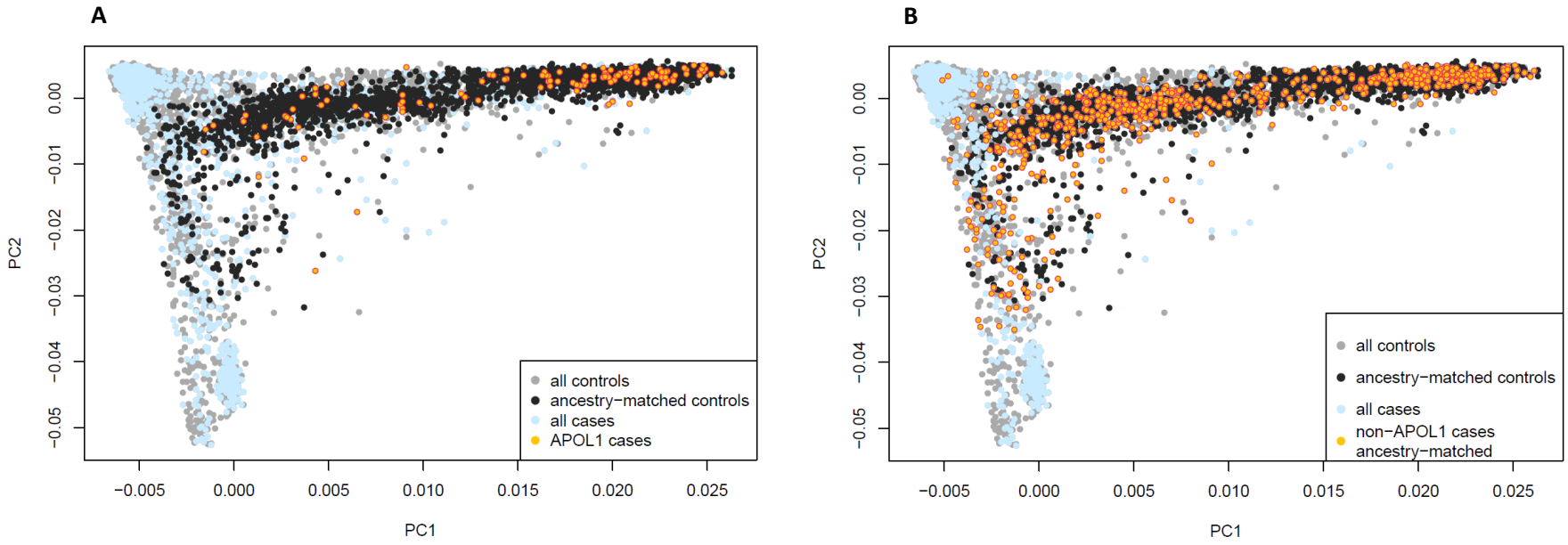
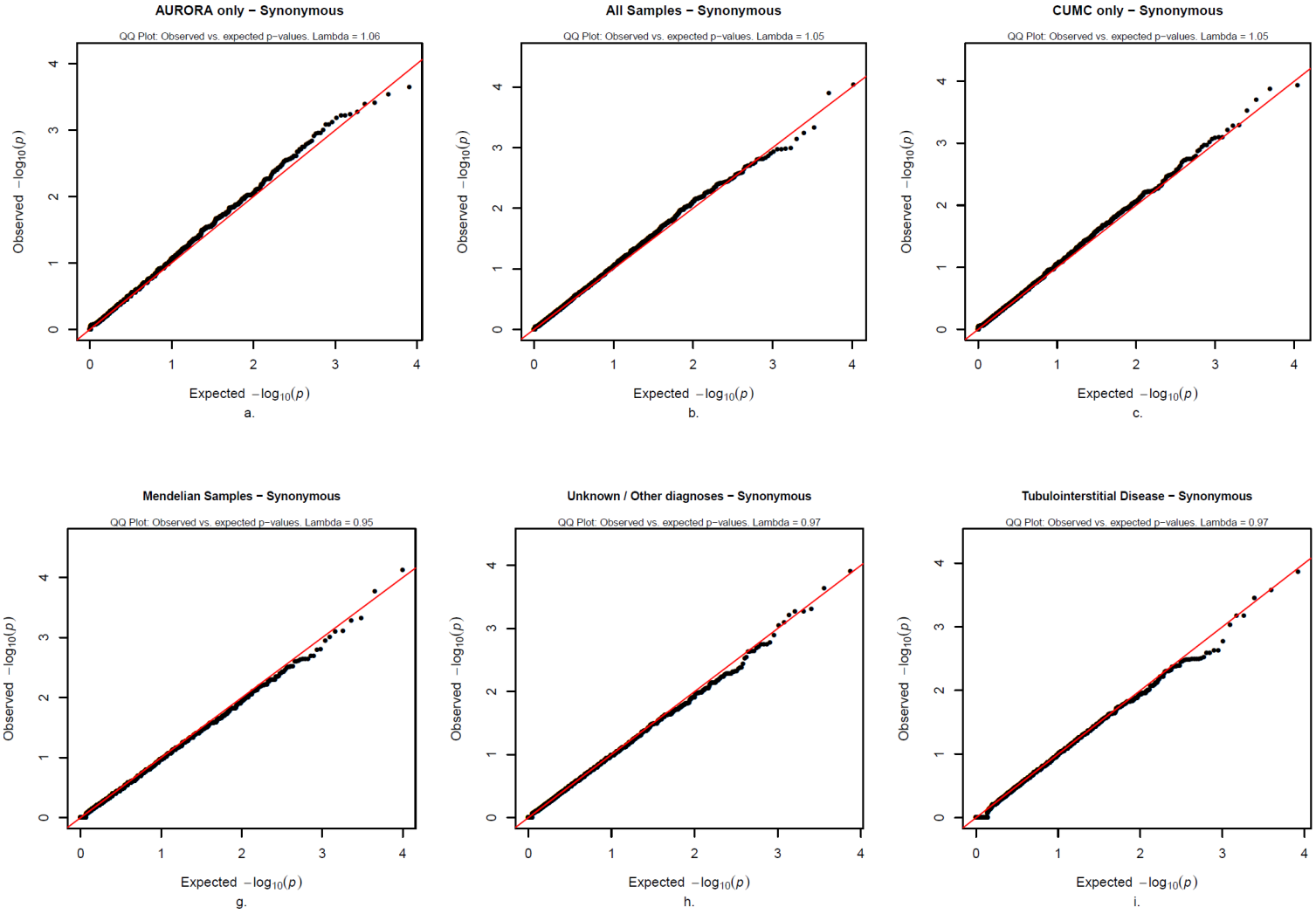
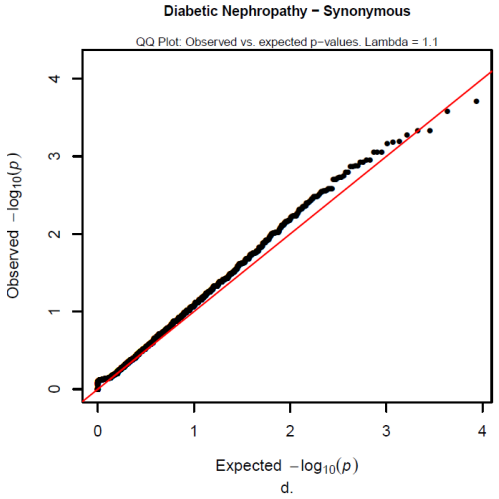
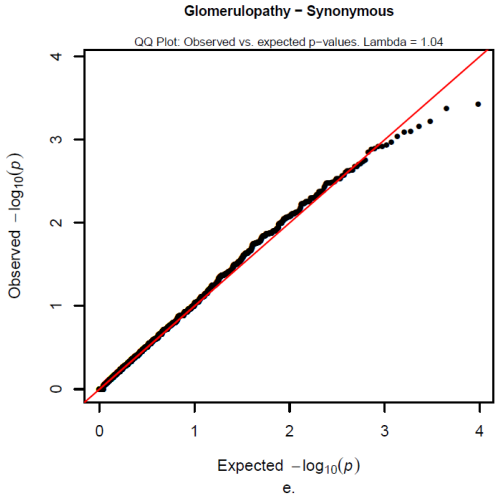


Figure S2. Quantile-Quantile plots for the dominant synonymous models of each subgroup. (continued on next page)

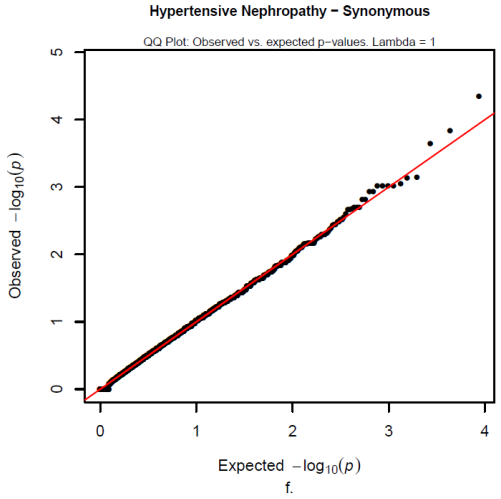



Diabetic Nephropathy – Synonymous
QQ Plot: Observed vs. expected p–values. Lambda = 1.1
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
0
1
2
3
4

d.

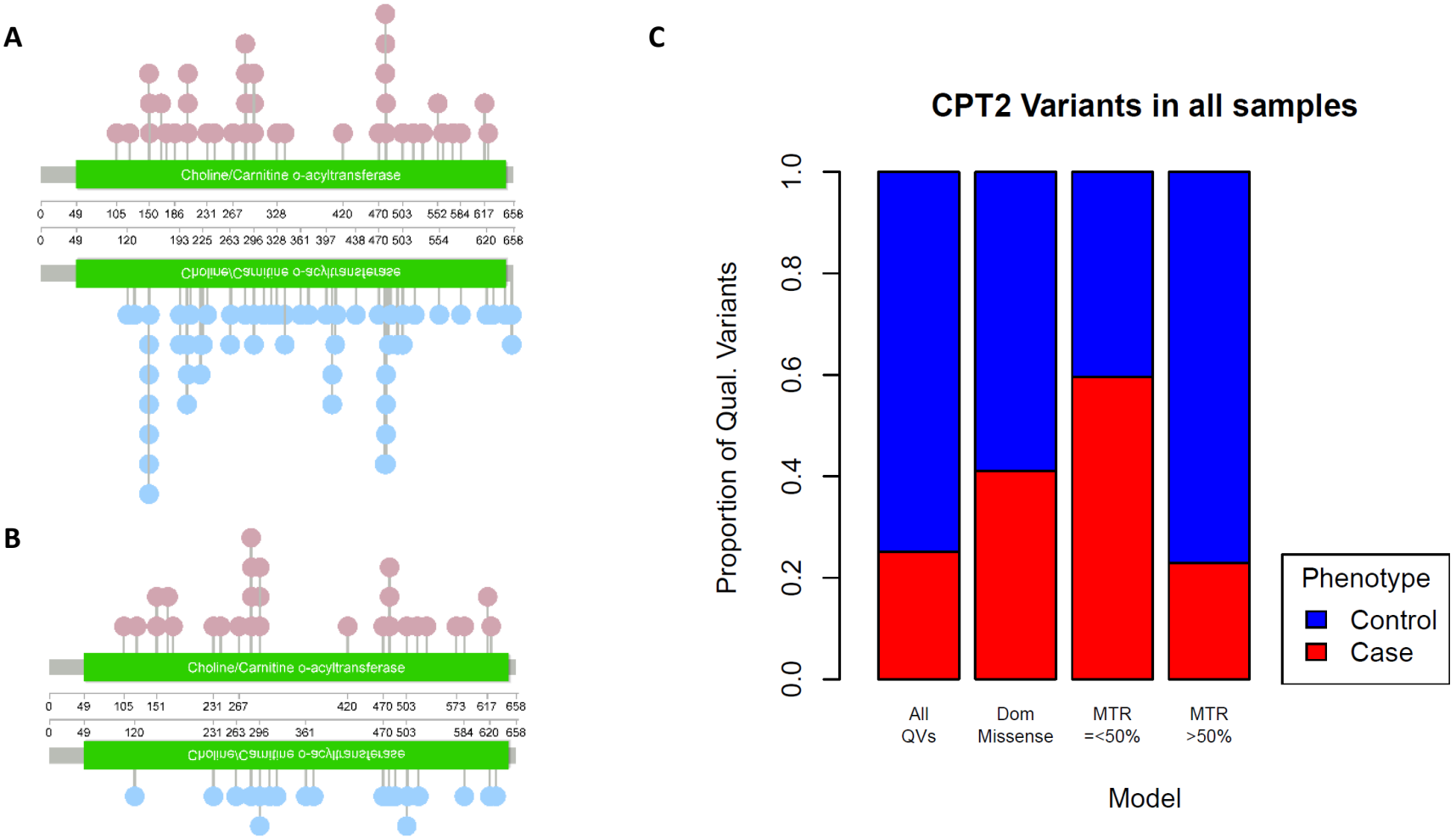

Glomerulopathy – Synonymous
QQ Plot: Observed vs. expected p–values. Lambda = 1.04
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
0
1
2
3
4

e.


Hypertensive Nephropathy – Synonymous
QQ Plot: Observed vs. expected p–values. Lambda = 1
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
0
1
2
3
4
5

f.

Supplemental Figure S3: Lollipop diagram of rare missense variants in *CPT2*, across case and control collections, with no intolerance filter (A) and after qualifying the MTR ≤ 50th percentile threshold (B). Case variant distribution is plotted in red and control is plotted in blue. The proportion of control and case variants across different categories of rare missense variants is shown in (C) and highlights the signal detection improvement driven by adopting the MTR ≤ 50th percentile threshold (p = 7.45x10^{-8}, 95% CI lower = 0.44, 95% CI higher = 0.74, exact binomial test where observed proportion of case variants in MTR≤50th % = 0.60, expected proportion from MTR>50th % = 0.23).



Supplemental Figure S4: Power curve for different control carrier frequencies.

Illustrating the number of cases required to achieve study-wide significance under a two-sided Fisher's exact test ($\alpha = 2.7x10^{-8}$), with 80% power and assuming a constant cohort growth at the current 25%:75% (case:control) ratio. The estimated number of cases required to reach the study-wide significance threshold is represented along the y-axis. The different control carrier frequency scenarios are illustrated by coloured curves.

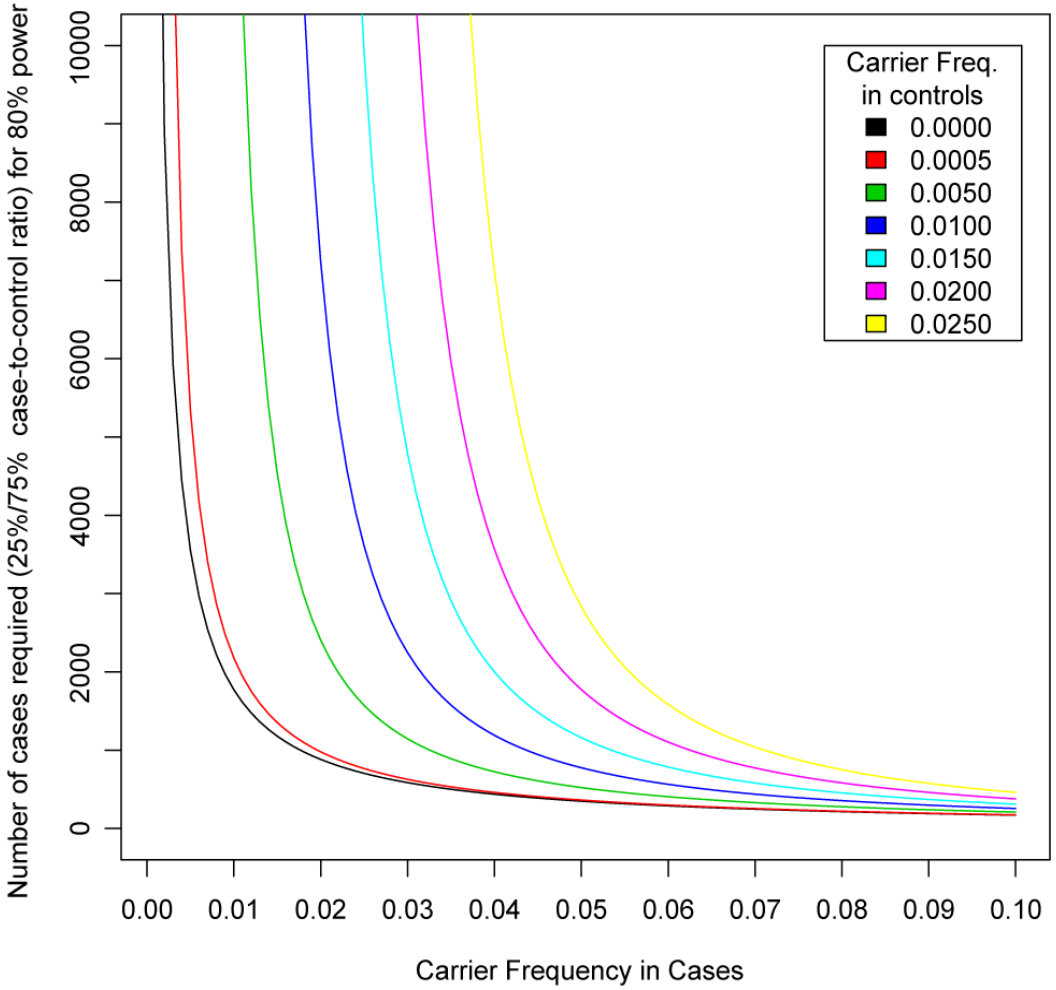


Table S1: Cohort Demographics

	AURORA		CUMC		Control		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
	Gender							
Male	685	61.8%	1160	56.8%	3881	40.6%	5726	45.0%
Female	424	38.2%	881	43.2%	5682	59.4%	6987	55.0%
	Genotype predicted ancestry							
African	15	1.4%	394	19.3%	1887	19.7%	2296	18.1%
Caucasian	981	88.5%	797	39.0%	4749	49.7%	6527	51.3%
East Asian	6	0.5%	120	5.9%	124	1.3%	250	2.0%
Hispanic	56	5.0%	365	17.9%	1536	16.1%	1957	15.4%
Middle Eastern	2	0.2%	40	2.0%	179	1.9%	221	1.7%
South Asian	4	0.4%	17	0.8%	21	0.2%	42	0.3%
Unknown	45	4.1%	308	15.1%	1067	11.2%	1420	11.2%
	Age							
Mean (SD)	64.4 (8.6)		44.6 (20)		NA		NA	

Genetic ancestry for CKD samples was estimated by using PCA results, as previously described.[10]

Table S2: Control cohort phenotypes

	Controls	
Broad diagnosis	Number	Percent
Healthy (no disease ascertainment)	6437	67.3%
Dementia	710	7.4%
Common Epilepsies*	2377	24.9%
Other Neurological	39	0.4%
Total	**9563**	**100.0%**

*No paediatric epilepsies included.

Table S3: Qualifying Variant parameters and subgroups. The internal cohort is the cohort being studied (AURORA, CUMC and controls).

	Function	External MAF (gnomAD*)	MAF (internal cohort)	Polyphen-2 (HumVar) Filter	Tolerance filter	Tested Groups
#Ultra-rare deleterious predicted (D)	PTV + missense + in-frame indel	0	0.05%	Probably	None	9 tests; Combined, CUMC, AURORA, 6 clinical subgroups
Ultra-rare non-benign (D)	PTV + missense + in-frame indel	0	0.05%	Probably, Possibly, Unknown	None	9 tests; Combined, CUMC, AURORA, 6 clinical subgroups
#Rare protein-truncating (D)	PTV	0.1%	0.10%	NA	None	9 tests; Combined, CUMC, AURORA, 6 clinical subgroups
#Rare non-benign (D)	PTV + missense + in-frame indel	0.05%	0.10%	Probably, Possibly, Unknown	None	9 tests; Combined, CUMC, AURORA, 6 clinical subgroups
Rare missense only (D)	Missense	0.05%	0.10%	Probably, Possibly, Unknown, Benign	None	9 tests; Combined, CUMC, AURORA, 6 clinical subgroups
#Rec non-syn (R)	PTV + missense + in-frame indel	1.00%	1.00%	Probably, Possibly, Unknown, Benign	None	1 test; Combined (n=1,778 European ancestry cases)
Ultra-rare non-benign MTR50 (D)	PTV + missense + in-frame indel	0	0.05%	Probably, Possibly, Unknown	MTR ≤50%	9 tests; Combined, CUMC, AURORA, 6 clinical subgroups
Ultra-rare non-benign LIMBR50 (D)	PTV + missense + in-frame indel	0	0.05%	Probably, Possibly, Unknown	LIMBR ≤50%	9 tests; Combined, CUMC, AURORA, 6 clinical subgroups
Rare non-benign MTR50 (D)	PTV + missense + in-frame indel	0.05%	0.10%	Probably, Possibly, Unknown	MTR ≤50%	9 tests; Combined, CUMC, AURORA, 6 clinical subgroups
Rare non-benign LIMBR50 (D)	PTV + missense + in-frame indel	0.05%	0.10%	Probably, Possibly, Unknown	LIMBR ≤50%	9 tests; Combined, CUMC, AURORA, 6 clinical subgroups
Rare missense only MTR50 (D)	Missense	0.05%	0.10%	Probably, Possibly, Unknown, Benign	MTR ≤50%	9 tests; Combined, CUMC, AURORA, 6 clinical subgroups
Rare missense only LIMBR50 (D)	Missense	0.05%	0.10%	Probably, Possibly, Unknown, Benign	LIMBR ≤50%	9 tests; Combined, CUMC, AURORA, 6 clinical subgroups
#Rare syn (D)	Synonymous	0.05%	0.10%	NA	None	9 tests; Combined, CUMC, AURORA, 6 clinical subgroups
Rec syn (R)	Synonymous	1.00%	1.00%	NA	None	1 test; Combined (n=1,778 European ancestry cases)

*gnomAD max minor allele frequency (MAF) imposed across the six genetic ancestries African, American, Non-Finnish European, Finnish, East Asian, Ashkenazi Jewish and South Asian population MAFs. #These models are consistent with our previous collapsing analysis publications.[11, 12] PTV = Putative Protein Truncating Variant (incl. start lost, nonsense, frameshift, canonical splice), indel = insertion/deletion; Non-benign includes start lost, nonsense, frameshift, canonical splice, missense and inframe indels with the PolyPhen filter applying to missense variants, MTR = Missense Tolerance Ratio within gene percentile, LIMBR = localized intolerance model using Bayesian regression percentile. D = Dominant model, R/Rec = Recessive model (incl. homozygous, hemizygous and a compound heterozygous proxy whereby two qualifying variants identified in gene in absence of phase information).

Table S4: LIMBR and MTR comparison to standard missense-only collapsing analysis run

Gene Name	Subgroup	Model name	Exome-wide rank	Qualified case	Unqualified case	Qualified case freq.	Qualified ctrl	Unqualified ctrl	Qualified ctrl freq.	Enriched direction	Odds ratio	P	CI 95 lower	CI 95 upper
COL4A3	i-all_AURORA	(Dominant) Rare Missense	316	35	1074	0.0316	193	9370	0.0202	case	1.58	0.0204	1.10	2.28
COL4A3	i-all_AURORA	(Dominant) Rare Missense LIMBR50	9598	0	1109	0	5	9558	5.23E-04	ctrl	0.00	1	0.00	14.78
COL4A3	i-all_AURORA	(Dominant) Rare Missense MTR50	1242	18	1091	0.0162	101	9462	0.0106	case	1.55	0.0956	0.93	2.56
COL4A3	v-AURORA-CUMC-all	(Dominant) Rare Missense	676	83	3067	0.0263	192	9371	0.0201	case	1.32	0.0404	1.02	1.71
COL4A3	v-AURORA-CUMC-all	(Dominant) Rare Missense LIMBR50	15176	1	3149	3.18E-04	5	9558	5.23E-04	ctrl	0.61	1	0.07	5.20
COL4A3	v-AURORA-CUMC-all	(Dominant) Rare Missense MTR50	3062	42	3108	0.0133	101	9462	0.0106	case	1.27	0.2057	0.88	1.82
COL4A3	viii-CUMC-all	(Dominant) Rare Missense	3067	50	1991	0.0245	192	9371	0.0201	case	1.23	0.2012	0.89	1.68
COL4A3	viii-CUMC-all	(Dominant) Rare Missense LIMBR50	10097	1	2040	4.90E-04	5	9558	5.23E-04	ctrl	0.94	1	0.11	8.03
COL4A3	viii-CUMC-all	(Dominant) Rare Missense MTR50	6954	25	2016	0.0122	101	9462	0.0106	case	1.16	0.4815	0.75	1.80
COL4A4	i-all_AURORA	(Dominant) Rare Missense	12825	21	1088	0.0189	175	9388	0.0183	case	1.04	0.814	0.66	1.64
COL4A4	i-all_AURORA	(Dominant) Rare Missense LIMBR50	14899	0	1109	0	0	9563	0	NA	NA	1	NA	NA
COL4A4	i-all_AURORA	(Dominant) Rare Missense MTR50	2524	6	1103	0.0054	94	9469	0.0098	ctrl	0.55	0.1862	0.24	1.25
COL4A4	v-AURORA-CUMC-all	(Dominant) Rare Missense	14527	60	3090	0.019	179	9384	0.0187	case	1.02	0.8801	0.76	1.37
COL4A4	v-AURORA-CUMC-all	(Dominant) Rare Missense LIMBR50	15175	0	3150	0	0	9563	0	NA	NA	1	NA	NA
COL4A4	v-AURORA-CUMC-all	(Dominant) Rare Missense MTR50	6925	26	3124	0.0083	95	9468	0.0099	ctrl	0.83	0.4592	0.54	1.28
COL4A4	viii-CUMC-all	(Dominant) Rare Missense	13899	39	2002	0.0191	176	9387	0.0184	case	1.04	0.8564	0.73	1.47
COL4A4	viii-CUMC-all	(Dominant) Rare Missense LIMBR50	15067	0	2041	0	0	9563	0	NA	NA	1	NA	NA
COL4A4	viii-CUMC-all	(Dominant) Rare Missense MTR50	14394	20	2021	0.0098	94	9469	0.0098	ctrl	1.00	1	0.61	1.62
COL4A5	i-all_AURORA	(Dominant) Rare Missense	1281	20	1089	0.018	111	9452	0.0116	case	1.56	0.0817	0.97	2.53
COL4A5	i-all_AURORA	(Dominant) Rare Missense LIMBR50	14901	0	1109	0	1	9562	1.05E-04	ctrl	0.00	1	0.00	138.07
COL4A5	i-all_AURORA	(Dominant) Rare Missense MTR50	17	17	1092	0.0153	56	9507	0.0059	case	2.64	0.0014	1.53	4.56
COL4A5	v-AURORA-CUMC-all	(Dominant) Rare Missense	24	62	3088	0.0197	110	9453	0.0115	case	1.73	9.40E-04	1.26	2.36

COL4A5	v-AURORA-CUMC-all	(Dominant) Rare Missense LIMBR50	10603	0	3150	0	1	9562	1.05E-04	ctrl	0.00	1	0.00	48.56
COL4A5	v-AURORA-CUMC-all	(Dominant) Rare Missense MTR50	7	40	3110	0.0127	55	9508	0.0058	case	2.22	1.90E-04	1.48	3.35
COL4A5	viii-CUMC-all	(Dominant) Rare Missense	29	42	1999	0.0206	110	9453	0.0115	case	1.81	0.0018	1.26	2.58
COL4A5	viii-CUMC-all	(Dominant) Rare Missense LIMBR50	15069	0	2041	0	0	9563	0	NA	NA	1	NA	NA
COL4A5	viii-CUMC-all	(Dominant) Rare Missense MTR50	122	23	2018	0.0113	55	9508	0.0058	case	1.97	0.0102	1.21	3.21
CPT2	i-all_AURORA	(Dominant) Rare Missense	8	18	1091	0.0162	56	9507	0.0059	case	2.80	7.31E-04	1.64	4.78
CPT2	i-all_AURORA	(Dominant) Rare Missense LIMBR50	8	18	1091	0.0162	56	9507	0.0059	case	2.80	7.31E-04	1.64	4.78
CPT2	i-all_AURORA	(Dominant) Rare Missense MTR50	1	15	1094	0.0135	19	9544	0.002	case	6.89	4.46E-07	3.49	13.59
CPT2	v-AURORA-CUMC-all	(Dominant) Rare Missense	11	39	3111	0.0124	56	9507	0.0059	case	2.13	4.71E-04	1.41	3.21
CPT2	v-AURORA-CUMC-all	(Dominant) Rare Missense LIMBR50	9	39	3111	0.0124	56	9507	0.0059	case	2.13	4.71E-04	1.41	3.21
CPT2	v-AURORA-CUMC-all	(Dominant) Rare Missense MTR50	1	28	3122	0.0089	19	9544	0.002	case	4.51	4.05E-07	2.51	8.08
CPT2	viii-CUMC-all	(Dominant) Rare Missense	511	21	2020	0.0103	56	9507	0.0059	case	1.76	0.0341	1.07	2.92
CPT2	viii-CUMC-all	(Dominant) Rare Missense LIMBR50	331	21	2020	0.0103	56	9507	0.0059	case	1.76	0.0341	1.07	2.92
CPT2	viii-CUMC-all	(Dominant) Rare Missense MTR50	29	13	2028	0.0064	19	9544	0.002	case	3.22	0.0019	1.59	6.53
PKD1	i-all_AURORA	(Dominant) Rare Missense	6	107	1002	0.0965	639	8924	0.0668	case	1.49	4.73E-04	1.20	1.85
PKD1	i-all_AURORA	(Dominant) Rare Missense LIMBR50	1	100	1009	0.0902	575	8988	0.0601	case	1.55	1.95E-04	1.24	1.93
PKD1	i-all_AURORA	(Dominant) Rare Missense MTR50	2	70	1039	0.0631	306	9257	0.032	case	2.04	1.06E-06	1.56	2.66
PKD1	v-AURORA-CUMC-all	(Dominant) Rare Missense	309	251	2899	0.0797	640	8923	0.0669	case	1.21	0.0158	1.04	1.41
PKD1	v-AURORA-CUMC-all	(Dominant) Rare Missense LIMBR50	102	232	2918	0.0737	576	8987	0.0602	case	1.24	0.009	1.06	1.45
PKD1	v-AURORA-CUMC-all	(Dominant) Rare Missense MTR50	14	143	3007	0.0454	304	9259	0.0318	case	1.45	4.35E-04	1.18	1.77
PKD1	viii-CUMC-all	(Dominant) Rare Missense	9321	143	1898	0.0701	640	8923	0.0669	case	1.05	0.5932	0.87	1.27
PKD1	viii-CUMC-all	(Dominant) Rare Missense LIMBR50	5600	131	1910	0.0642	576	8987	0.0602	case	1.07	0.5075	0.88	1.30
PKD1	viii-CUMC-all	(Dominant) Rare Missense MTR50	5997	73	1968	0.0358	309	9254	0.0323	case	1.11	0.4129	0.86	1.44
PKD2	i-all_AURORA	(Dominant) Rare Missense	11602	11	1098	0.0099	87	9476	0.0091	case	1.09	0.7395	0.58	2.05
PKD2	i-all_AURORA	(Dominant) Rare Missense LIMBR50	2761	8	1101	0.0072	46	9517	0.0048	case	1.50	0.2643	0.71	3.19
PKD2	i-all_AURORA	(Dominant) Rare Missense MTR50	11528	5	1104	0.0045	40	9523	0.0042	case	1.08	0.8064	0.42	2.74
PKD2	v-AURORA-CUMC-all	(Dominant) Rare Missense	12282	27	3123	0.0086	90	9473	0.0094	ctrl	0.91	0.7472	0.59	1.40

PKD2	v-AURORA-CUMC-all	(Dominant) Rare Missense LIMBR50	10726	16	3134	0.0051	49	9514	0.0051	ctrl	0.99	1	0.56	1.75
PKD2	v-AURORA-CUMC-all	(Dominant) Rare Missense MTR50	13654	15	3135	0.0048	43	9520	0.0045	case	1.06	0.8789	0.59	1.91
PKD2	viii-CUMC-all	(Dominant) Rare Missense	8129	15	2026	0.0073	88	9475	0.0092	ctrl	0.80	0.5153	0.46	1.38
PKD2	viii-CUMC-all	(Dominant) Rare Missense LIMBR50	5193	7	2034	0.0034	47	9516	0.0049	ctrl	0.70	0.474	0.31	1.54
PKD2	viii-CUMC-all	(Dominant) Rare Missense MTR50	12900	9	2032	0.0044	41	9522	0.0043	case	1.03	0.8541	0.50	2.12
SCLT1	i-all_AURORA	(Dominant) Rare Missense	7434	3	1106	0.0027	47	9516	0.0049	ctrl	0.55	0.4821	0.17	1.77
SCLT1	i-all_AURORA	(Dominant) Rare Missense LIMBR50	14950	0	1109	0	0	9563	0	NA	NA	1	NA	NA
SCLT1	i-all_AURORA	(Dominant) Rare Missense MTR50	4700	1	1108	9.02E-04	26	9537	0.0027	ctrl	0.33	0.3555	0.04	2.44
SCLT1	v-AURORA-CUMC-all	(Dominant) Rare Missense	5048	20	3130	0.0063	46	9517	0.0048	case	1.32	0.3168	0.78	2.24
SCLT1	v-AURORA-CUMC-all	(Dominant) Rare Missense LIMBR50	15216	0	3150	0	0	9563	0	NA	NA	1	NA	NA
SCLT1	v-AURORA-CUMC-all	(Dominant) Rare Missense MTR50	15251	8	3142	0.0025	26	9537	0.0027	ctrl	0.93	1	0.42	2.06
SCLT1	viii-CUMC-all	(Dominant) Rare Missense	1008	17	2024	0.0083	46	9517	0.0048	case	1.74	0.0657	0.99	3.04
SCLT1	viii-CUMC-all	(Dominant) Rare Missense LIMBR50	15113	0	2041	0	0	9563	0	NA	NA	1	NA	NA
SCLT1	viii-CUMC-all	(Dominant) Rare Missense MTR50	9699	7	2034	0.0034	26	9537	0.0027	case	1.26	0.6453	0.55	2.91

Table S5: All collapsing analysis results for all genes across all models, where P < 0.05. This table is available as a separate file (SUPP_TABLE_S5_Final.xlsx)

Column descriptors for Table S5

Rank.Within.Model – Gene rank (by P value) with the corresponding subgroup analysis

model.name – subgroup collapsing model (Table S3)

Gene.Name – gene name (HGNC symbol)

Enriched.Direction – whether enrichment of qualified variants are in cases or controls

Fet.P – P-value (Fisher's Exact Test)

OR – odds ratio (Fisher's Exact Test)

known.ckd.gene – whether a human renal phenotype with Mendelian inheritance has been previously reported for mutations in this gene (1 = yes, 0 = no)

Qualified.Case – number of cases carrying a qualifying variant

Unqualified.Case – number of cases not carrying a qualifying variant

Qualified.Case.Freq – frequency of cases carrying a qualifying variant

Qualified.Ctrl – number of controls carrying a qualifying variant

Unqualified.Ctrl – number of controls not carrying a qualifying variant

Qualified.Ctrl.Freq – frequency of controls carrying a qualifying variant

RVIS.ExAC. – Exac-based Residual Variation Intolerance Score [13] for the gene. Lower scores indicate the gene is more intolerant to non-synonymous variation.

GenicConstraint_mis.z.ExAC. – ExAC-based Constraint (missense z) scores [14] for the gene. Lower scores indicate greater constraint on non-synonymous mutations.

LoF.FDR.ExAC. – An exome-wide FDR adjusted loss-of-function depletion p-value [15] generated to find genes that are specifically depleted of LoF genetic variation using the ExAC reference cohort.

LoF.pLI.ExAC. – ExAC-based probability of being loss-of-function intolerant (intolerant of both heterozygous and homozygous LoF variants).

LoF.pRec.ExAC. – ExAC-based probability of being intolerant of homozygous, but not heterozygous LoF variants.

LoF.pNull.ExAC. – ExAC-based probability of being tolerant of both heterozygous and homozygous LoF variants.

Supplemental table S6. Distribution of phenotypes in cases with qualifying variants in *PKD1, PKD2, COL4A3, COL4A4 and COL4A5*

Collapsing Analysis Results						Clinical Diagnosis					
Gene	**group**	**model**	**Fisher exact P-value**	**Odds ratio**	**Total qualified cases**	**Diabetes Count (%)**	**GN Count (%)**	**Hypertension Count (%)**	**Mendelian / Congenital Anomaly Count (%)**	**Tubulo-interstitial Count (%)**	**Unknown/other Count(%)**
PKD1	All samples	dom_ultrarare_OO	1.66E-22	6	83	3 (3.61%)	9 (10.8%)	2 (2.4%)	64 (77.1%)	1 (1.2%)	4 (4.8%)
PKD2	All samples	dom_rare_LOF	7.27E-13	61	20	0	0	0	20 (100%)	0	0
COL4A5	All samples	dom_rare_mtr50	4.80E-07	2.8	49	4 (8.2%)	13 (26.5%)	5 (10.2%)	18 (36.7%)	3 (6.1%)	6 (12.7%)
COL4A4	All samples	dom_ultrarare_LIMBR50*	1.01E-05	8.5	14	0	4 (28.6%)	3 (21.4%)	5 (35.7%)	2 (15.3%)	0
COL4A3	All samples	dom_ultrarare_OO	3.05E-05	7.9	13	0	6 (46.1%)	1 (7.7%)	3 (23.1%)	2 (15.4%)	1 (7.7%)

***All missense variants were excluded in this gene under the LIMBR50 filter, and all qualifying variants (case and control) are predicted Loss of Function**

Table S7. Ancestry of samples in the modifier analysis, using PCA of genotypes.

	PKD1/2		COL4A3/4/5		APOL1 high risk	
Population	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
African	3	3.19%	11	12.60%	109	79.60%
Caucasian	84	89.40%	51	58.60%	0	0.00%
East Asian	0	0.00%	2	2.30%	0	0.00%
Hispanic*	1	1.06%	11	12.60%	27	19.70%
Middle Eastern	1	1.06%	1	1.15%	0	0.00%
South Asian	0	0.00%	0	0.00%	0	0.00%
Unknown^	5	5.32%	11	12.60%	1	0.73%

*Note that those who self-identified as 'Hispanic' ethnicity in the comparison group were predominantly of Afro-Caribbean descent.

^ 'Unknown' was defined as a multinomial prediction $p < 0.95$ for all listed ancestry groups.

Table S8: AHDC1 rare QV rates in cases with APOL1 genotypes (dual risk, heterozygous, and homozygous reference) against ancestry-matched controls

	Qualified Case	Unqual-ified Case	Qualified Ctrl	Unqual-ified Ctrl	Odds Ratio	P	CI 95 lower	CI 95 higher
AHDC1 in dual risk APOL1 (hom or comp het G1 / G2)	9	128	11	2198	14	5.91E-07	5.03	37.91
AHDC1 in Het G1 or G2	2	172	11	2198	2.322	0.24	0.25	10.77
AHDC1 in Hom Ref G1 and G2	4	427	11	2198	1.871	0.29	0.43	6.35

Table S9: Association results of top genes from the collapsing analysis with CKD phenotypes in the UK biobank (GeneATLAS database)

		Gene	*AHDC1*	*CPT2*	*COL4A3*	*COL4A4*	*PKD2*	*SCLT1*	*SLC17A1*	*PKD1*	*COL4A5*
		Chr	1	1	2	2	4	4	6	16	X
		Gene Start	27873815	53662616	228029443	227872041	88928886	129805632	25799013	2139728	107683356
		Gene Stop	27878626	53679267	228176586	228012199	88996846	130014258	25830785	2185690	107939608
		Lowest P	6.49E-05	9.35E-05	1.97E-05	2.51E-17	7.54E-05	7.71E-06	1.87E-05	4.16E-06	1.96E-05
		Lowest P SNP	rs147012676	rs144844459	rs188942711	rs35138315	rs1481012	rs190446297	rs183313411	rs200930978	rs1028404
Trait Key	**Trait Description**										
clinical_c_Block_N00-N08	N00-N08 Glomerular diseases	n.nom.snps	0	0	0	1	0	0	0	1	0
		lowest.p	NA	NA	NA	2.08E-09	NA	NA	NA	4.16E-06	NA
clinical_c_Block_N10-N16	N10-N16 Renal tubulo-interstitial diseases	n.nom.snps	0	0	0	0	0	4	0	0	0
		lowest.p	NA	NA	NA	NA	NA	7.71E-06	NA	NA	NA
clinical_c_Block_N20-N23	N20-N23 Urolithiasis	n.nom.snps	0	0	0	0	6	0	0	3	2
		lowest.p	NA	NA	NA	NA	8.36E-06	NA	NA	5.24E-05	8.18E-05
clinical_c_Block_O10-O16	O10-O16 Oedema, proteinuria and hypertensive disorders in pregnancy, childbirth and the puerperium	n.nom.snps	0	0	0	0	0	0	0	1	0
		lowest.p	NA	NA	NA	NA	NA	NA	NA	6.91E-05	NA
clinical_c_N03	N03 Chronic nephritic syndrome	n.nom.snps	0	1	0	1	0	0	0	0	0
		lowest.p	NA	9.35E-05	NA	1.66E-05	NA	NA	NA	NA	NA
clinical_c_N18	N18 Chronic renal failure	n.nom.snps	0	0	0	1	0	0	0	0	0
		lowest.p	NA	NA	NA	6.23E-08	NA	NA	NA	NA	NA
clinical_c_N19	N19 Unspecified renal failure	n.nom.snps	1	0	1	0	0	0	0	0	0
		lowest.p	6.49E-05	NA	1.97E-05	NA	NA	NA	NA	NA	NA
clinical_c_N20	N20 Calculus of kidney and ureter	n.nom.snps	0	0	0	0	1	0	20	0	8
		lowest.p	NA	NA	NA	NA	7.54E-05	NA	4.30E-05	NA	1.96E-05
clinical_c_N23	N23 Unspecified renal colic	n.nom.snps	0	0	1	0	0	0	0	0	0
		lowest.p	NA	NA	9.00E-05	NA	NA	NA	NA	NA	NA
selfReported_n_1213	renal/kidney failure	n.nom.snps	0	0	0	0	0	0	2	0	0
		lowest.p	NA	NA	NA	NA	NA	NA	1.87E-05	NA	NA
selfReported_n_1218	kidney stone/ureter stone/bladder stone	n.nom.snps	0	0	0	0	0	0	12	0	0
		lowest.p	NA	NA	NA	NA	NA	NA	4.16E-05	NA	NA
selfReported_n_1454	other renal/kidney problem	n.nom.snps	0	0	0	1	0	1	0	0	0
		lowest.p	NA	NA	NA	2.51E-17	NA	7.72E-06	NA	NA	NA

SNPs within 50Kb upstream/downstream of each gene were interrogated in GeneATLAS data (http://geneatlas.roslin.ed.ac.uk/, last accessed 2018-12-01). The lowest p-value for each trait are indicated, and the lowest SNP overall for each gene across all traits. Traits with p-values >1x 10-4 are indicated as NA.

Table S10: Predicting possible p-value signals among an increasing sized test cohort.

Gene	Case Frequency (n=3150)	Control Frequency (n=9563)	Fisher's Exact (FET) P-value	FET 2-fold	FET (3-fold case) (2-fold control)	FET 3-fold	Case L95%	Case U95%	Control L95%	Control U95%
PKD1	0.0263	0.0045	1.66E-22	2.17E-43	2.33E-54	3.27E-64	0.02104	0.03256	0.003256	0.006052
PKD2	0.0038	0.0001	5.29E-07	5.25E-13	1.55E-15	5.85E-19	0.00197	0.006645	2.65E-06	5.82E-04
COL4A3	0.0041	0.0005	3.05E-05	2.86E-09	2.93E-11	3.07E-13	0.002199	0.007047	0.00017	0.00122
TUBA1C	0.0022	0	5.71E-05	3.25E-09	7.97E-11	1.85E-13	0.000894	0.004573	0	0.000386
COL4A4	0.0051	0.001	7.52E-05	2.07E-08	3.18E-10	6.51E-12	0.002906	0.008235	0.000502	0.001922
CDH24	0.0032	0.0003	1.15E-04	3.43E-08	7.21E-10	1.16E-11	0.001523	0.00583	6.47E-05	9.17E-04
EDEM2	0.0032	0.0003	1.15E-04	3.43E-08	7.21E-10	1.16E-11	0.001523	0.00583	6.47E-05	9.17E-04
ANXA2	0.0029	0.0002	1.16E-04	3.07E-08	7.03E-10	9.18E-12	0.001307	0.005417	2.53E-05	7.55E-04
COL4A5	0.0048	0.0012	3.46E-04	4.30E-07	1.29E-08	8.30E-10	0.002668	0.007842	0.000574	0.002057
NTNG1	0.0022	0.0001	3.58E-04	2.31E-07	9.46E-09	1.67E-10	0.000894	0.004573	2.65E-06	5.82E-04
SIPA1	0.0022	0.0001	3.58E-04	2.31E-07	9.46E-09	1.67E-10	0.000894	0.004573	2.65E-06	5.82E-04
HSPA1L	0.0032	0.0005	7.31E-04	1.54E-06	6.78E-08	3.70E-09	0.001523	0.00583	0.00017	0.00122
SDR42E1	0.0016	0	9.32E-04	8.68E-07	6.14E-08	8.08E-10	0.000516	0.0037	0	0.000386
MB21D2	0.0016	0	9.32E-04	8.68E-07	6.14E-08	8.08E-10	0.000516	0.0037	0	0.000386
IMMT	0.0016	0	9.32E-04	8.68E-07	6.14E-08	8.08E-10	0.000516	0.0037	0	0.000386
ZNF439	0.0025	0.0003	0.0011	3.04E-06	1.74E-07	9.49E-09	0.001097	0.004998	6.47E-05	9.17E-04
AR	0.0041	0.001	0.0011	6.29E-06	2.33E-07	2.13E-08	0.002199	0.007047	0.000502	0.001922
FAM189A1	0.0035	0.0007	0.0011	4.02E-06	3.61E-07	1.62E-08	0.001744	0.00624	0.000294	0.001508
NOX3	0.0035	0.0007	0.0011	4.02E-06	3.61E-07	1.62E-08	0.001744	0.00624	0.000294	0.001508
VPS33B	0.0022	0.0002	0.0013	3.49E-06	2.22E-07	1.09E-08	0.000894	0.004573	2.53E-05	7.55E-04
SLC12A6	0.0022	0.0002	0.0013	3.49E-06	2.22E-07	1.09E-08	0.000894	0.004573	2.53E-05	7.55E-04
SMYD1	0.0022	0.0002	0.0013	3.49E-06	2.22E-07	1.09E-08	0.000894	0.004573	2.53E-05	7.55E-04
CDNF	0.0019	0.0001	0.0013	2.88E-06	1.99E-07	7.31E-09	0.000699	0.004141	2.65E-06	5.82E-04
NOS2	0.0029	0.0005	0.002	1.12E-05	1.14E-06	7.14E-08	0.001307	0.005417	0.00017	0.00122
CDH2	0.0029	0.0005	0.002	1.12E-05	1.14E-06	7.14E-08	0.001307	0.005417	0.00017	0.00122
GPR128	0.0025	0.0004	0.0026	1.77E-05	1.44E-06	1.37E-07	0.001097	0.004998	0.000114	0.001071
MRC2	0.0025	0.0004	0.0026	1.77E-05	1.44E-06	1.37E-07	0.001097	0.004998	0.000114	0.001071
SPN	0.0025	0.0004	0.0026	1.77E-05	1.44E-06	1.37E-07	0.001097	0.004998	0.000114	0.001071
ACAD11	0.0025	0.0004	0.0026	1.77E-05	1.44E-06	1.37E-07	0.001097	0.004998	0.000114	0.001071
CCDC147	0.0025	0.0004	0.0026	1.77E-05	1.44E-06	1.37E-07	0.001097	0.004998	0.000114	0.001071

Reflecting on the top 30 genes in the all-cause CKD ultra-rare deleterious model we provide examples of possible enrichment p-value signals as the cohort size scales up (two- and/or three-fold) under the strict assumption that the current case and control carrier frequency estimates remain true and that the case composition remains the same during the up-scale. Orange shaded cells indicate genes that surpass experiment-wide significance threshold of $p < 2.7x10^{-8}$.

FET = Fisher's Exact test.

FET 2-fold = A Fisher's Exact Test p-value assuming the currently observed case and control frequencies remain the same among double the cases (n=6,300) and double the controls (n=19,126).

FET (3-fold case) (2-fold control) = A Fisher's Exact Test p-value assuming the currently observed case and control frequencies remain the same among triple the cases (n=9,450) and double the controls (n=19,126).

FET 3-fold = A Fisher's Exact Test p-value assuming the currently observed case and control frequencies remain the same among triple the cases (n=9,450) and triple the controls (n=28,689).

Case L95%/U95% = The 95%CI sampling variability range for the current observed case carrier frequency estimate.

Control L95%/U95% = The 95%CI sampling variability range for the current observed control carrier frequency estimate.

References:

1. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010;26(22):2867-73.
2. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics. 2006;38(8):904-9.
3. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-91. doi: 10.1038/nature19057. PubMed PMID: 27535533; PubMed Central PMCID: PMC5018207.
4. Petrovski S, Todd JL, Durheim MT, Wang Q, Chien JW, Kelly FL, et al. An Exome Sequencing Study to Assess the Role of Rare Genetic Variation in Pulmonary Fibrosis. American journal of respiratory and critical care medicine. 2017;(ja).
5. Gussow AB, Petrovski S, Wang Q, Allen AS, Goldstein DB. The intolerance to functional genetic variation of protein domains predicts the localization of pathogenic mutations within genes. Genome Biol. 2016;17:9. doi: 10.1186/s13059-016-0869-4. PubMed PMID: 26781712; PubMed Central PMCID: PMCPMC4717634.
6. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Research. 2017;45(Database issue):D200-D3. doi: 10.1093/nar/gkw1129. PubMed PMID: PMC5210587.
7. Hayeck TJ, Stong N, Wolock CJ, Copeland B, Kamalakaran S, Goldstein DB, et al. Improved Pathogenic Variant Localization via a Hierarchical Model of Sub-regional Intolerance. Am J Hum Genet. 2019;104(2):299-309. Epub 2019/01/29. doi: 10.1016/j.ajhg.2018.12.020. PubMed PMID: 30686509; PubMed Central PMCID: PMCPMC6369453.
8. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nature methods. 2010;7(4):248-9.
9. Gelfman SD, S., Araujo Martins Moreno C, Ren Z, Wolock C, Shneider N, Phatnani H, et al. Regional collapsing of rare variation implicates specific regions in known and novel ALS genes. In Prep. 2018.
10. Petrovski S, Goldstein DB. Unequal representation of genetic variation across ancestry groups creates healthcare inequality in the application of precision medicine. Genome Biol. 2016;17(1):157. doi: 10.1186/s13059-016-1016-y. PubMed PMID: 27418169; PubMed Central PMCID: PMCPMC4944427.

11. Petrovski S, Todd JL, Durheim MT, Wang Q, Chien JW, Kelly FL, et al. An Exome Sequencing Study to Assess the Role of Rare Genetic Variation in Pulmonary Fibrosis. Am J Respir Crit Care Med. 2017;196(1):82-93. doi: 10.1164/rccm.201610-2088OC. PubMed PMID: 28099038; PubMed Central PMCID: PMCPMC5519963.
12. Epi4K consortium. Ultra-rare genetic variation in common epilepsies: a case-control sequencing study. The Lancet Neurology. 2017;16(2):135-43. Epub 2017/01/20. doi: 10.1016/s1474-4422(16)30359-3. PubMed PMID: 28102150.
13. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes. PLOS Genetics. 2013;9(8):e1003709. doi: 10.1371/journal.pgen.1003709.
14. Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, et al. A framework for the interpretation of de novo mutation in human disease. Nature Genetics. 2014;46:944. doi: 10.1038/ng.3050

https://www.nature.com/articles/ng.3050#supplementary-information.
15. Petrovski S, Gussow AB, Wang Q, Halvorsen M, Han Y, Weir WH, et al. The Intolerance of Regulatory Sequence to Genetic Variation Predicts Gene Dosage Sensitivity. PLoS genetics. 2015;11(9):e1005492. doi: 10.1371/journal.pgen.1005492. PubMed PMID: 26332131; PubMed Central PMCID: PMC4557908.